IN THE SPECIFICATION:

(i) Please amend the paragraph on page 12, line 20, through page 13, line12, as follows:

For instance, in one embodiment, the data type identification module 14 may identify data types by examining the geometric patterns of the input data and comparing the geometric patterns with known patterns that characterize particular data types. For example, typed textual data is characterized by features such as high level symmetry of lines, different characters have parallel strokes, sharp angles, and are relatively the same height, etc. On the other hand, handwritten characters are ~~characterize~~ characterized by variations in size and direction of strokes, etc. In addition, the prototype database 15 preferably comprises different types of fonts for typed textual characters, wherein the identification module 14 can compare input textual data with the textual data in database 15 to find a matching font (using scaling and a suitable distance measure) to thereby define the data type and font of the input data.

(ii) Please amend the paragraph on page 22, line 23, through page 24, line 5, as follows:

The counter 33 outputs the semantic units and corresponding counts 40 and phonetic syllable 41. The semantic units 40 represent a character string that bears some semantic meaning (e.g., syllables or morphemes such as roots in Russian words which are not actual words, but represent a common semantic meaning for different words that contain the root). The phonetic syllables 41 comprise a special string of character that represent how some strings of characters (corresponding to the given semantic units), e.g., syllables, sound. The phonetic syllables 41 and semantic units 40 (e.g., syllables /morphemes) are used by a language model generator 42 to derive probabilities of distribution of phonetic syllables given syllable 41 and generate a language model based on semantic units. In particular, using techniques known in the art, the syllable counts and conditional distributions of phonetic syllables 41 are used to construct LM of phonetic syllables. For example, this procedure is similar to constructing a language model for classes such as described in the articles by Eugene Charniak, entitled "Statistical Language Learning", The MIT Press, Cambridge, 1996; and Frederick Jelinek, "Statistical Methods for Speech Recognition", The MIT Press, Cambridge, 1998.) Methods for generating a language model for morphemes, for example, are described in U.S. Patent No. 6,073,091, which issued on June 6, ~~200~~ 2000 to Kanevsky et al., entitled "Apparatus and Method For Forming A Filtered Inflected Language Model for Automatic Speech Recognition" and U.S. Patent No. 5,835,888, which issued on November 10, 1998 to Kanevsky, et al., entitled "Statistical Language Model For Inflected Languages," both of which are fully incorporated herein by reference.